Bioinformatics, 33(14), 2017, 2197–2198
doi: 10.1093/bioinformatics/btx144
Advance Access Publication Date: 16 March 2017
Applications Note

OXFORD

Genome analysis

SynMap2 and SynMap3D: web-based whole-genome synteny browsers

Asher Haug-Baltzell,[1] Sean A. Stephens,[2] Sean Davey,[3] Carlos E. Scheidegger[2] and Eric Lyons[1,3,*]

[1]Genetics GIDP, [2]Department of Computer Science and [3]BIO5 Institute, School of Plant Sciences, University of Arizona, Tucson AZ 85721, USA

*To whom correspondence should be addressed.
Associate Editor: John Hancock

Received on September 26, 2016; revised on December 21, 2016; editorial decision on March 10, 2017; accepted on March 14, 2017

Abstract

Summary: Current synteny visualization tools either focus on small regions of sequence and do not illustrate genome-wide trends, or are complicated to use and create visualizations that are difficult to interpret. To address this challenge, The Comparative Genomics Platform (CoGe) has developed two web-based tools to visualize synteny across whole genomes. SynMap2 and SynMap3D allow researchers to explore whole genome synteny patterns (across two or three genomes, respectively) in responsive, web-based visualization and virtual reality environments. Both tools have access to the extensive CoGe genome database (containing over 30 000 genomes) as well as the option for users to upload their own data. By leveraging modern web technologies there is no installation required, making the tools widely accessible and easy to use.

Availability and Implementation: Both tools are open source (MIT license) and freely available for use online through CoGe (https://genomevolution.org). SynMap2 and SynMap3D can be accessed at http://genomevolution.org/coge/SynMap.pl and http://genomevolution.org/coge/SynMap3D.pl, respectively. Source code is available: https://github.com/LyonsLab/coge.

Contact: ericlyons@email.arizona.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Two or more genomic regions are considered 'syntenic' if they are derived from an ancestral genomic region. Syntenic regions can be inferred by identifying colinear sets of homologous genes. Locating and comparing syntenic regions among multiple genetic sequences, both inter- and intraspecies, is common practice in many comparative genomic studies to understand genomic conservation and evolution. With a current database of over 30 000 genomes for 17 000 species and the underlying infrastructure for users to upload and store public or private genomes, the comparative genomics platform CoGe is uniquely positioned to provide synteny analysis and visualization systems to researchers (Lyons and Freeling, 2008). Here we describe two new web-based whole-genome synteny visualization tools that use dot-plots to visualize synteny. SynMap2 is a high-performance two-genome synteny viewer, and SynMap3D is a novel three-genome synteny browser. These systems are important steps for increasing researchers' ability to visualize and interact with the rapidly growing set of sequenced genomes.

2 Description

Currently available synteny visualization tools (i.e. Krzywinski *et al.*, 2009; McKay *et al.*, 2010) are useful, but are limited by a lack of interactivity and complicated setup/use. Further, most current systems focus on small regions of genomic sequence and do not represent genome-wide trends. Those that do permit whole-genome comparisons create visualizations that can be difficult to interpret, especially with regards to polyploidy. To address these challenges,

CoGe has developed two new web-based tools that complement its existing tools for whole genome and micro-synteny analyses, SynMap2 and SynMap3D. Both tools visualize synteny across whole genomes using dot-plots, and illustrate mutation rate patterns through color-mapped histograms. SynMap2 (Supplementary Fig. S1A) is a dynamic, interactive, high-performance web-based visualizer for comparing two whole genomes, and boasts many new features and performance benefits over the original SynMap (Lyons *et al.*, 2008). SynMap3D (Supplementary Fig. 1B) is a novel, 3D synteny viewer that identifies synteny common among three genomes.

2.1 SynMap2

SynMap2 is a major reworking of the original SynMap pairwise whole-genome synteny viewer. It provides a number of benefits over the original SynMap tool by replacing static images with a high performance and interactive web-based dot-plot viewer. SynMap2's interactive features allow researchers to explore the dataset from multiple angles without requiring additional knowledge of database query languages. Subsets of syntenic matches can be selected by chromosomal regions or by subset ranges of mutation values (Kn, Ks or Kn/Ks). As subsets are selected, the visualization is automatically updated. The subset selection algorithms are designed to work efficiently and are built upon the open-source project Crossfilter (http://square.github.io/crossfilter/), a fast data structure for incremental data querying and aggregation.

SynMap2's dot-plots are visualized as 2D scatter plots. Visuals are generated through a combination of D3 (https://d3js.org/), SVG (Dahlström *et al.*, 2011), and a custom implementation of the data points using HTML5 Canvas (Cabanier *et al.*, 2015). This custom implementation avoids the overhead of keeping individual data points in the Document Object Model. Dots are colored based on their mutation ratios using a data-driven colormap designed to split the histogram of mutation values into clusters, such that clusters of similar values are given similar colors. A technique from computational topology known as persistence-based simplification is used to identify the most important peaks in the histogram to generate these colormaps (Edelsbrunner and Harer, 2010). Coordinating the colors between the histogram and dot-plot illustrates genome-wide patterns (e.g. polyploidy) and highlights genomic regions with different drift ratios and evolutionary origins (e.g. lineage divergence). A number of example SynMap2 analyses can be found in Supplementary Table S1.

2.2 SynMap3D

SynMap3D is a novel addition to the CoGe toolbox designed to compare and visualize whole-genome synteny across three genomes in an interactive, three dimensional dot-plot. SynMap3D allows researchers to specify three genomes, and uses CoGe's computational pipelines to dynamically identify synteny gene sets among all three genomes. The results are displayed to the user in the form of a 3D scatter plot. Visualizations are rendered through the open-source technologies D3 and Three.js (http://threejs.org/), with modifications to enhance performance and provide intuitive interaction with the data. Interactive features and color-mapped histogram of mutation values is similar to SynMap2. SynMap3D also includes the option to enable virtual reality (VR) mode, allowing researchers with a VR headset to interact with their data in an immersive, responsive environment. To take advantage of VR mode, users must have a VR-compatible browser and VR-headset (e.g. Oculus Rift). A number of example SynMap3D analyses can be found in Supplementary Table S2.

3 Availability and requirements

Both tools are available online through CoGe (http://www.genomevolution.org). Both require a modern HTML5 capable browser, and SynMap3D requires WebGL be enabled. VR mode in SynMap3D requires a VR-compatible browser (Mozilla Firefox or Google Chrome) and a VR headset. Users can visualize any of CoGe's publically available genomes without creating a user account, but must register to upload their own genomes or access restricted data. More detailed use instructions, help, and example analyses for are available on each tools respective CoGePedia page (https://genomevolution.org/wiki/index.php/SynMap3D and https://genomevolution.org/wiki/index.php/SynMap2).

Acknowledgements

We would like to thank CyVerse (formerly iPlant) for sharing best practices and providing data and computational infrastructure for CoGe (NSF DBI—1265383), and the contributors to D3 and Three.js their roles in developing and testing powerful web visualization frameworks.

Funding

This work has been supported by the National Science Foundation grants [IOS—1339156 and IOS—1444490].

Conflict of Interest: none declared.

References

Cabanier,R. *et al.* (2015) *HTML Canvas 2D Context*. World Wide Web Consortium.

Dahlström,E. *et al.* (2011) *Scalable Vector Graphics (SVG) 1.1*, 2nd edn. World Wide Web Consortium.

Edelsbrunner,H. and Harer,J. (2010) Computational topology: an introduction. *Am. Math. Soc.*

Krzywinski,M.I. *et al.* (2009) Circos: An information aesthetic for comparative genomics. *Genome Res.*, **19**, 1639–1645.

Lyons,E. and Freeling,M. (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J.*, **53**, 661–673.

Lyons,E. *et al.* (2008) The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. *Trop. Plant Biol.*, **1**, 181–190.

McKay,S.J. *et al.* (2010) Using the Generic Synteny Browser (GBrowse_syn). *Curr. Protoc. Bioinformatics*, Chapter 9: Unit 9.12.